



Queen's Economics Department Working Paper No. 1240

Numerical Distribution Functions of Fractional Unit Root and Cointegration Tests

James G. MacKinnon
Queen's University

Morten Ørregaard Nielsen
Queen's University and CREATES

Department of Economics
Queen's University
94 University Avenue
Kingston, Ontario, Canada
K7L 3N6

7-2010 (revised 5-2012)

Numerical distribution functions of fractional unit root and cointegration tests*

James G. MacKinnon[†]
Queen's University

Morten Ørregaard Nielsen
Queen's University and CREATES

August 26, 2012

Abstract

We calculate, by simulations, numerical asymptotic distribution functions of likelihood ratio tests for fractional unit roots and cointegration rank. Because these distributions depend on a real-valued parameter b which must be estimated, simple tabulation is not feasible. Partly due to the presence of this parameter, the choice of model specification for the response surface regressions used to obtain the numerical distribution functions is more involved than is usually the case. We deal with model uncertainty by model averaging rather than by model selection. We make available a computer program which, given the dimension of the problem, q , and a value of b , provides either a set of critical values or the asymptotic P value for any value of the likelihood ratio statistic.

JEL Codes: C12, C16, C22, C32

Keywords: cofractional process, cointegration rank, fractional unit root, fractional cointegration, likelihood ratio test, model averaging, numerical CDF, response surface regression

1 Introduction

This paper calculates, by simulation, numerical asymptotic distribution functions – and hence critical values and P values – of tests for fractional unit roots and cointegration rank based on the fractional vector autoregressive (FVAR) model proposed in Johansen (2008) and analyzed by Johansen and Nielsen (2010, 2012), henceforth JN. For a time series \mathbf{X}_t of dimension p , the FVAR model is given in error correction form as

$$\Delta^d \mathbf{X}_t = \Delta^{d-b} L_b \boldsymbol{\alpha} \boldsymbol{\beta}' \mathbf{X}_t + \sum_{i=1}^k \boldsymbol{\Gamma}_i \Delta^d L_b^i \mathbf{X}_t + \boldsymbol{\varepsilon}_t, \quad t = 1, \dots, T, \quad (1)$$

*We are grateful to the editor, Hashem Pesaran, Bruce Hansen, Søren Johansen, and three anonymous referees for comments and discussion, and to Russell Davidson for the use of his computer cluster. We would also like to thank participants at the Colloquium on Computationally-Intensive Econometrics at McGill University, May, 2009. This research was supported by the Social Sciences and Humanities Research Council of Canada (SSHRC grants 410-2009-0183 and 410-2009-0194) and the Center for Research in Econometric Analysis of Time Series (CREATES, funded by the Danish National Research Foundation).

[†]Corresponding author. Address: Department of Economics, Queen's University, Dunning Hall, 94 University Avenue, Kingston, ON K7L 3N6, Canada. Email: jgm@econ.queensu.ca

where $\boldsymbol{\varepsilon}_t$ is p -dimensional *i.i.d.* $(\mathbf{0}, \boldsymbol{\Omega})$, $d \geq b > 0$, Δ^d is the fractional difference operator, and $L_b = 1 - \Delta^b$ is the fractional lag operator. For the model with $d = b$ it is also possible to include a constant term to obtain

$$\Delta^d \mathbf{X}_t = \boldsymbol{\alpha}(\boldsymbol{\beta}' L_d \mathbf{X}_t + \boldsymbol{\rho}') + \sum_{i=1}^k \boldsymbol{\Gamma}_i \Delta^d L_d^i \mathbf{X}_t + \boldsymbol{\varepsilon}_t, \quad t = 1, \dots, T, \quad (2)$$

noting that $L_d \boldsymbol{\rho}' = \boldsymbol{\rho}'$ because $\Delta^d 1 = 0$.

The analysis in JN (2012) is based on the conditional probability measure and conditional likelihood. Thus, the observations $\mathbf{X}_1, \dots, \mathbf{X}_T$ are generated by the models (1) and (2) as functions of parameters, innovations $\boldsymbol{\varepsilon}_1, \dots, \boldsymbol{\varepsilon}_T$, and initial values $\mathbf{X}_{-n}, n \geq 0$, and the latter are assumed fixed and bounded. For the calculations of the conditional likelihood function and the maximum likelihood estimator, the fractional difference $\Delta^a \mathbf{X}_t$ is needed for different values of a , and this depends on the infinitely many initial values $\mathbf{X}_{-n}, n \geq 0$, which are not all observed. It is assumed that observations on \mathbf{X}_t for $t = -N_0 + 1, \dots, T$ are available. Although the infinitely many $\mathbf{X}_{-n}, n \geq 0$, enter in the data generating process, the calculations of the conditional likelihood and associated estimators and test statistics are based on $\tilde{\mathbf{X}}_{-n} = \mathbf{X}_{-n}, n = 0, \dots, N_0 - 1$, and $\tilde{\mathbf{X}}_{-n} = 0, n \geq N_0$. Thus N_0 observations are set aside for initial values, as is usually done in the analysis of an AR(N_0) model.

The models (1) and (2) include the Johansen (1988, 1991) cointegrated VAR model as the special case $d = b = 1$, and the interpretation of the model parameters is similar. Thus, we can write $\boldsymbol{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}'$, where the $p \times r$ matrices $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, with $r \leq p$, are assumed to have full column rank r . Then the columns of $\boldsymbol{\beta}$ are the r cointegrating (cofractional) relations that determine the long-run equilibria, and $\boldsymbol{\alpha}$ contains the adjustment coefficients. The parameter $\boldsymbol{\rho}$ is the so-called restricted constant term (since the constant term in the model is restricted to be of the form $\boldsymbol{\mu} = \boldsymbol{\alpha}\boldsymbol{\rho}'$), such that $\boldsymbol{\beta}' \mathbf{X}_t + \boldsymbol{\rho}'$ has mean zero. The rank r is termed the cointegrating, or cofractional, rank. The expansion of $L_b = 1 - \Delta^b$ has no term in L^0 , and thus only lagged disequilibrium errors appear in (1) and (2).

JN (2012) shows how to estimate the parameters of the models (1) and (2), including the fractional orders d and b , and how to test hypotheses on the cointegration rank r . The tests include, as special cases, Johansen's (1988, 1991) tests for cointegration rank in VARs (when $d = b = 1$ is known), standard Dickey-Fuller tests (when $d = b = 1$ is known and $p = 1$), and fractional Dickey-Fuller tests (JN, 2010) (when $p = 1$). A Matlab package is available for calculating estimators and test statistics in models (1) and (2); see Nielsen and Morin (2012).

Programs for computing accurate asymptotic critical values and asymptotic P values, based on numerical distribution functions (NDFs), are available for standard Dickey-Fuller tests (MacKinnon, 1996) and Johansen tests (MacKinnon, Haug, and Michelis, 1999). Our objective is to obtain comparable NDFs for cointegration rank tests in models (1) and (2).

In fractionally integrated models, a new complication is the dependence of the asymptotic distributions on the real-valued parameter b . This implies that critical values have to be simulated conditional on an estimate of this parameter. An additional complication that arises partially because of the presence of the parameter b is that the choice of model specification for the response surface regressions is more involved than is the case for ordinary unit root and cointegration tests. One methodological innovation of this paper is that we use model averaging rather than model selection to deal with model uncertainty in the

specification of response surface regressions.

The remainder of the paper is organized as follows. Section 2 briefly introduces the tests and associated asymptotic distributions. Section 3 discusses how we obtain numerical approximations to the distribution functions of the tests. Section 4 explains how our results can be used to obtain approximate asymptotic critical values and P values via computer programs that we make available, and Section 5 concludes.

2 Likelihood ratio tests for fractional cointegration

JN (2012) shows that asymptotic theory is standard when $b < 0.5$, and therefore we focus in this paper on the case $d \geq b > 0.5$, where asymptotic theory is non-standard and involves fractional Brownian motion. Specifically, when $b > 0.5$, JN (2012) shows that, under i.i.d. errors with suitable moment conditions, the maximum likelihood (assuming normality, and conditional on initial values) parameter estimates $\hat{d}, \hat{b}, \hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\Gamma}}_1, \dots, \hat{\boldsymbol{\Gamma}}_k$ are asymptotically Gaussian, while the estimates $\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\rho}}$ are locally asymptotically mixed normal. These results allow asymptotically standard (chi-squared) inference on all parameters of the model, including the cointegrating relations and orders of fractionality, using quasi-likelihood ratio tests.

In this paper, we are concerned with the asymptotic distribution of the likelihood ratio (LR) test statistic for the hypothesis $\mathcal{H}_r: \text{rank}(\boldsymbol{\Pi}) = r$ against $\mathcal{H}_p: \text{rank}(\boldsymbol{\Pi}) = p$ in model (1) and $\mathcal{H}_r: \text{rank}(\boldsymbol{\Pi}, \boldsymbol{\mu}) = r$ against $\mathcal{H}_p: \text{rank}(\boldsymbol{\Pi}, \boldsymbol{\mu}) = p$ in model (2). Let $L(d, b, r)$ be the conditional profile likelihood function under \mathcal{H}_r (with the restriction $d = b$ imposed in the case of model (2)), where $\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Gamma}_1, \dots, \boldsymbol{\Gamma}_k, \boldsymbol{\Omega}$ (and $\boldsymbol{\rho}$ in model (2)) have been concentrated out by regression and reduced rank regression; see JN (2012, Section 3.1) for details.

To calculate the LR statistic, we maximize the conditional profile likelihood function under the hypotheses \mathcal{H}_r and \mathcal{H}_p . Then $\text{LR}_T(q) = 2 \log(L(\hat{d}_p, \hat{b}_p, p)/L(\hat{d}_r, \hat{b}_r, r))$, where $q = p - r$,

$$L(\hat{d}_p, \hat{b}_p, p) = \max_{d,b} L(d, b, p), \text{ and } L(\hat{d}_r, \hat{b}_r, r) = \max_{d,b} L(d, b, r),$$

with the obvious modification for model (2) with $d = b$. The asymptotic distribution of this test statistic is derived in JN (2012), in which it is shown that, when $0.5 < b \leq d$,

$$\text{LR}_T(q) \xrightarrow{D} \text{Tr} \left\{ \int_0^1 d\mathbf{W}(s) \mathbf{F}'(s) \left(\int_0^1 \mathbf{F}(s) \mathbf{F}'(s) ds \right)^{-1} \int_0^1 \mathbf{F}(s) d\mathbf{W}'(s) \right\} = \text{LR}(q), \quad (3)$$

where the vector process $d\mathbf{W}$ is the increment of ordinary q -dimensional vector Brownian motion, with $q = p - r$. The vector process \mathbf{F} depends on the deterministic terms in a similar way as it does in the cointegrated VAR model in Johansen (1991). In model (1) where there is no deterministic term in the model, $\mathbf{F}(u) = \mathbf{W}_{b-1}(u)$. We refer to this as the “no constant” or NC case. In model (2) where the restricted constant term is included, $\mathbf{F}(u) = (\mathbf{W}'_{b-1}(u), 1)'$. We refer to this as the “restricted constant” or RC case. In both cases, $\mathbf{W}_{b-1}(u)$ is vector fractional Brownian motion of type II defined as

$$\mathbf{W}_{b-1}(u) = \boldsymbol{\Gamma}(b)^{-1} \int_0^u (u-s)^{b-1} d\mathbf{W}(s). \quad (4)$$

JN (2012, Section 2.3) shows that, in a version of Granger’s representation theorem, \mathbf{X}_t can be represented as

$$\mathbf{X}_t = \mathbf{C} \Delta^{-d} (\boldsymbol{\varepsilon}_t \mathbf{1}_{\{t \geq 1\}}) + \Delta^{-(d-b)} (\mathbf{Y}_t^+ \mathbf{1}_{\{t \geq 1\}}) + \boldsymbol{\mu}_t, \quad t = 1, \dots, T, \quad (5)$$

where $1_{\{A\}}$ is the indicator function for the event A , $\boldsymbol{\mu}_t$ depends on initial values of \mathbf{X}_t , and $\mathbf{Y}_t^+ = \sum_{n=0}^{t-1} \tau_n \boldsymbol{\varepsilon}_{t-n}$ is (asymptotically) a stationary process. JN (2012) shows that the asymptotic distributions do not depend on initial values under the assumption that these are either zero for $n \geq T^\nu$, $\nu < 0.5$, or satisfy $\sum_{n=1}^{\infty} n^{-1/2} \mathbf{X}_{-n} < \infty$. Given the representation (5), the functional central limit theorem for fractional processes shows that asymptotic distributions depend on type II stochastic integrals $\int_0^u (u-s)^{b-1} d\mathbf{W}(s)$ as in (3). These reflect the fact that there are no innovations prior to $t = 1$ because initial values are assumed fixed and inference is conditional on them, and thus only innovations for $t \geq 1$ are relevant for the limit distributions of test statistics and estimators. However, for the same reason, parameter estimates, test outcomes, etc. are, of course, conditional on initial values and may be different for different initial values. Obviously, such arguments are well known from conditional inference in AR models.

The asymptotic distribution $\text{LR}(q)$ of the test statistic $\text{LR}_T(q)$ defined in (3) depends on both $q = p - r$ and b . The dependence on the unknown real-valued parameter b complicates empirical analysis compared to the cointegrated VAR model. One approach would be to simulate the asymptotic distribution on a case-by-case basis, but this would be both impractical and imprecise, not to mention time-consuming. Another possibility would be to bootstrap the tests along the lines of a fractional version of Cavaliere, Rahbek, and Taylor (2012), although this would be extremely time-consuming due to the fractional differencing and numerical optimization required in each bootstrap replication. As a more practical alternative, we derive NDFs that can be used in practice to obtain accurate asymptotic critical values and P values for any dimension q and any given estimate of b .

3 Numerical distribution functions

In order to simulate $\text{LR}(q)$, we need to generate discrete-time analogs of \mathbf{W}_{b-1} and $d\mathbf{W}$. The latter is easy; we just use standard normal random q -vectors $\boldsymbol{\varepsilon}_t$. The former is a little harder. We discretize (4) and use $\mathbf{z}_{t-1} = T^{1/2-b} \boldsymbol{\Gamma}(b)^{-1} \sum_{s=1}^{t-1} (t-s)^{b-1} \boldsymbol{\varepsilon}_s$.¹ Note that \mathbf{z}_{t-1} depends on $\boldsymbol{\varepsilon}_s$ for $s < t$ but not on $\boldsymbol{\varepsilon}_t$. We further define the discretized version of \mathbf{F} as $\mathbf{s}_{t-1} = \mathbf{z}_{t-1}$ in model (1) where there is no deterministic term and $\mathbf{s}_{t-1} = (\mathbf{z}'_{t-1}, 1)'$ in model (2) where the restricted constant is included. Then the simulated version of $\text{LR}(q)$ in (3) is

$$\tau_q^b(T) = \text{Tr} \left\{ T^{-1/2} \sum_{t=2}^T \boldsymbol{\varepsilon}_t \mathbf{s}'_{t-1} \left(T^{-1} \sum_{t=2}^T \mathbf{s}_{t-1} \mathbf{s}'_{t-1} \right)^{-1} T^{-1/2} \sum_{t=2}^T \mathbf{s}_{t-1} \boldsymbol{\varepsilon}'_t \right\}, \quad (6)$$

which depends on q , T , and b .

Since we are interested in the asymptotic distribution $\text{LR}(q)$, $\tau_q^b(T)$ should ideally be simulated for $T = \infty$. Of course, this is impossible. Even using very large values of T is infeasible, because, for large T , the cost of simulation is roughly proportional to the square of T . The solution, as in MacKinnon (1996), is to simulate $\tau_q^b(T)$ for a large number of values of T , which range from quite small to moderately large, and use the results to estimate the asymptotic distribution by means of response surface regressions on estimated quantiles.

¹Alternative simulation strategies are discussed in Davidson and Hashimzade (2009), although it focuses on type I processes.

Most of the computational cost is incurred in generating the \mathbf{z}_t . To reduce this cost, we generate them for $q = 1, 2, \dots, 12$ simultaneously. For each value of T , we perform 200 experiments, and for each experiment we generate 1,199,999 realizations of (6) for $q = 1$, 599,999 realizations for $q = 2$, 399,999 realizations for $q = 3$, 299,999 realizations for $q = 4$, 199,999 realizations for $q = 5$ and $q = 6$, and 99,999 realizations for $7 \leq q \leq 12$. Thus the same random numbers enter into the calculations for all values of q . Of course, this scheme implies that our results will be most accurate for $q = 1$ and least accurate for $q \geq 7$.

The numbers of realizations always end in 99,999 so as to make the calculation of quantiles easy. For an experiment based on B realizations, the estimate of the π^{th} quantile is just number $\pi(B + 1)$ in the list of the $\tau_q^b(T)$ sorted from smallest to largest, provided that $\pi(B + 1)$ is an integer. For each of the 200 experiments for each value of T , we store 221 estimated quantiles (0.0001, 0.0002, 0.0005, 0.001, \dots , 0.010, 0.015, \dots , 0.990, 0.991, \dots , 0.999, 0.9995, 0.9998, 0.9999).

In principle, we are interested in all values of b in the half-open interval $0.5 < b \leq 2$. In practice, however, we must limit our simulations to a finite number of values and use an interpolation scheme. After some experimentation, we chose to perform the simulations for 31 values of b : 0.51, 0.55, 0.60, \dots , 1.95, 2.00.

For each value of q , b , and T , we obtain 200 estimates of each of the 221 quantiles. We then average these across the 200 experiments to obtain estimated quantiles $\hat{F}_q^b(T, \pi)$, where π takes on 221 values, along with standard errors $s_q^b(T, \pi)$. These estimated quantiles are the raw material from which we estimate NDFs by means of 221 response surface regressions for each value of b and q . These regressions use functions of T to explain the $\hat{F}_q^b(T, \pi)$.

Even though our experiments were carefully designed to minimize computational cost, they required many CPU-years of computer time. They also used so many random numbers that it was essential to employ a 64-bit random number generator. We actually combined two such generators, which were taken from L'Ecuyer *et al.* (1993).

In order to obtain asymptotic quantiles, we utilize response surface models that involve regressing the $\hat{F}_q^b(T, \pi)$ on a constant and powers of $T^{-b/2}$. This is in line with theoretical results in, e.g., Park (2003) who found that the next term in an asymptotic expansion of the Dickey-Fuller test statistic is of order $O_P(T^{-1/2})$. Hence, we use the models

$$\text{Models 1, 2, 3, 4 : } 1, T^{-b/2}, [T^{-b}], [T^{-3b/2}], [T^{-2b}].$$

Here square brackets indicate that a regressor does not appear in the lowest-numbered model but does appear in at least one higher-numbered model, with regressors being added sequentially from left to right as the model number increases. Thus Model 1 has just two regressors, namely, a constant and $T^{-b/2}$, while Model 4 has five regressors.

Although one or more of Models 1 through 4 often performs very well, that is not always the case. We therefore add five more models,

$$\text{Models 5, 6 : } 1, T^{-b/2}, T^{-b}, T^{-2b}, [T^{-3b}],$$

$$\text{Models 7, 8, 9 : } 1, T^{-b}, T^{-2b}, [T^{-3b}], [T^{-4b}].$$

We also experimented with a few models that had more than five regressors. However, these often showed evidence of severe overfitting.

What we are interested in for all the response surfaces is the constant term, which can be interpreted as an estimate of the π^{th} quantile of the asymptotic distribution. Ideally, all

models that fit well will lead to similar estimates of the constant term. As we explain below, that is true for the larger values of b but not always for the smaller ones.

The estimation method we use is weighted least squares. Following MacKinnon (1996), for every value of b , q , and π , we initially regress the square of $s_q^b(T, \pi)$ on a constant, $1/T$, and $1/T^2$ and use the inverse square roots of the fitted values from that regression as weights. We could have skipped the initial regression and simply used $1/s_q^b(T, \pi)$, but the smoothing that results from using fitted values makes the weights much less variable and should produce slightly more efficient estimates.

If the weights are asymptotically correct and the response surface regression accounts successfully for the systematic variation in the quantiles, the sum of squared residuals (SSR) from the weighted regression is asymptotically (in n , the number of values of T) distributed as $\chi^2(n-k)$, where k is the number of regressors. Thus it is very easy to tell when a response surface regression fits well and when it fits badly. In the former case, the SSR will be less than $n-k$ for more than half the values of π and less than the 0.95 quantile of the $\chi^2(n-k)$ distribution for about 95% of them. In the latter case, it will usually be a very big number.

In MacKinnon (1996) and other previous work on numerical distribution functions, model selection procedures were used to choose the best model for each set of 221 estimated quantiles. It is essential that these be the same for all π for a given set of experiments to prevent the resulting NDF from being non-smooth or even non-monotonic. However, model selection does not always work well, especially for smaller values of b . The chosen model, and consequently the NDF, sometimes changes substantially between adjacent values of b , even though the relative fits of alternative models change only modestly. Since we wish to use interpolation for values of b not directly covered by our experiments, it is very important that the NDFs vary smoothly as b changes. Instead of model selection, we therefore use model averaging; see Buckland *et al.* (1997) and Hansen (2007). To our knowledge, this is the first use of model averaging in this context.

Our criterion function for model j is

$$Q(j) = \frac{1}{221} \sum_{\pi} \left(\text{SSR}_j(\pi) + 2k_j \right),$$

where $\text{SSR}_j(\pi)$ denotes the SSR of model j for quantile π , k_j denotes the number of coefficients estimated for model j , and the summation over π means that we are summing over all 221 quantiles. The weight given to a particular model with $j = J$ when there are m models is then

$$W(J) = \frac{\exp(-Q(J))}{\sum_{j=1}^m \exp(-Q(j))}. \quad (7)$$

The formula (7) is based on one suggested by Buckland *et al.* (1997). It gives high weights to models that fit well and are parsimonious, and it ensures that the weights are nonnegative and sum to one. When the weight for any model is initially less than 10^{-4} , it is set to zero, and the remaining weights are rescaled so that they sum to one. This ensures that poorly fitting models have no effect on the results. The final estimate of each of the 221 asymptotic quantiles, for given b and q , is then

$$\hat{F}_q^b(\infty, \pi) = \sum_{j=1}^m W(j) \hat{\gamma}_{\pi}(j),$$

where $\hat{\gamma}_\pi(j)$ denotes the estimated constant term in the response surface regression for quantile π based on model j .

How many values of T we use, and which ones, varies somewhat across the experiments. The following 33 values are always used:

$$100, 110, \dots, 290, 300, 350, \dots, 550, 600, 700, \dots, 1200.$$

This choice produces satisfactory results for the largest values of b , but not for smaller values. We therefore add some additional, larger values of T for some cases. For $1.15 \leq b \leq 1.50$, we add 1300 and 1400. For $1.05 \leq b \leq 1.10$, we also add 1500 and 1600. For $b \leq 1.00$, we further add 1700, 1800, 1900, and 2000. Thus $n = 33, 35, 37$, or 41 . For the largest values of T , each set of 200 experiments took up to a week of computer time.

Larger values of T are needed as b becomes smaller because the shape of the relationship between \hat{F}_q^b and T changes with b . Experimental results for small values of T also become less informative about the asymptotic distribution as b becomes smaller. Both of these results are evident in Figure 1, which graphs $\hat{F}_q^b(T, 0.95)$, the estimated 0.05 critical value, against T for several values of b for the NC case. In the top panel, $q = 1$, and in the bottom panel, $q = 12$. Figure 2 contains results for the RC case that are comparable to the ones in Figure 1. Both the levels and the slopes of the $\hat{F}_q^b(T, 0.95)$ as functions of T look very different in this case.

The response surface regressions have a much easier time fitting the data for large values of b than for small ones. In the former case, all response surfaces that fit well tend to yield very similar estimates of $F_q^b(\infty, \pi)$. Thus, for large b , our NDFs suffer from very little model uncertainty. In the latter case, however, several response surfaces that all fit well can yield somewhat different estimates of $F_q^b(\infty, \pi)$. In consequence, despite the very large number of simulations that we performed, the NDFs for small values of b may suffer from some model uncertainty.

To make the above remarks more concrete, consider the 0.95 quantile (the 0.05 critical value) for $q = 1$ in the NC case. When $b = 1.50$, six models get weights greater than 0.02. Among these six, the smallest estimated asymptotic quantile is 4.5501, the largest is 4.5506, and the weighted average is 4.5504. Thus all the response surface regressions yield what is for all practical purposes the same estimate, and the weighted average is almost certainly very accurate. In contrast, when $b = 0.51$, five models get weights greater than 0.02. Among these five, the smallest estimated asymptotic quantile is 3.6161, the largest is 3.6866, and the weighted average is 3.6482. These estimates vary somewhat. With the model averaging approach, we avoid making an arbitrary choice among the models that all fit quite well.

It does not make sense to worry much about the possible inaccuracy of our NDFs for small values of b . As can be seen from the above example (the case $b = 0.51$ is the most extreme one) and from Figure 1, any such inaccuracy is small relative to the errors that must inevitably arise from using asymptotic rather than finite-sample distributions and from using estimates of b rather than true values.

The curves graphed in Figures 1 and 2 appear to be very smooth, even though these are simply raw results that have not been smoothed in any way. The curves appear to be smooth because they are based on a great many simulations, and there is consequently very little experimental error. It is because of this that poorly specified response surfaces produce very large values of $\text{SSR}_j(\pi)$.

Figure 1: Estimated 0.95 quantiles as functions of T , NC case

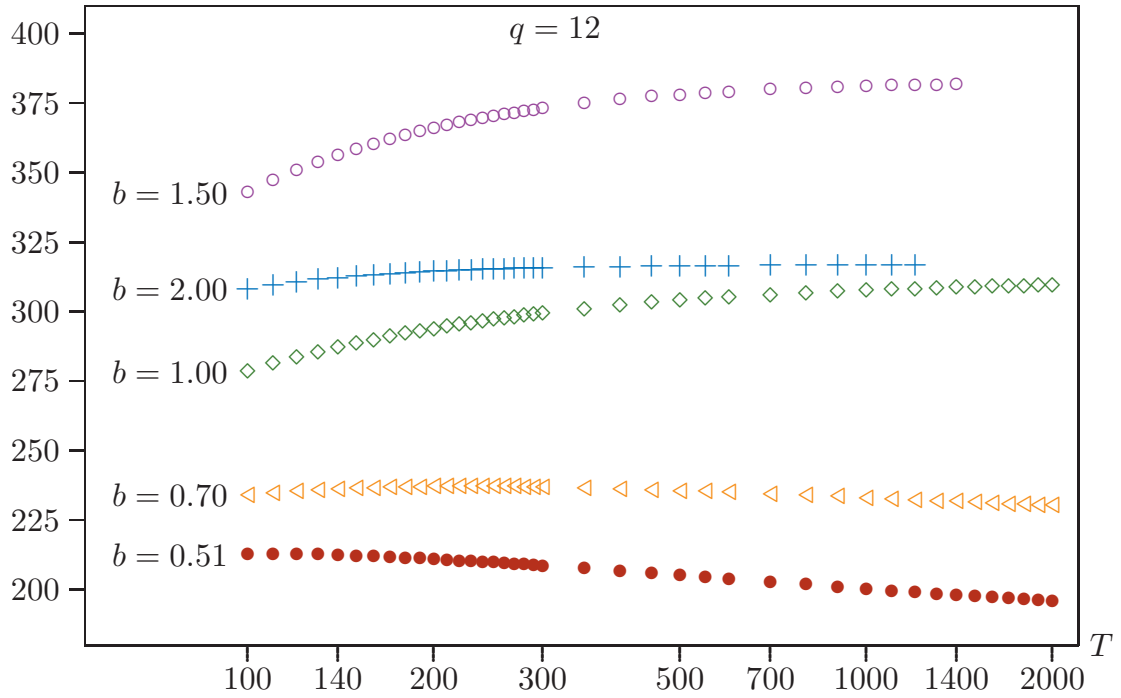
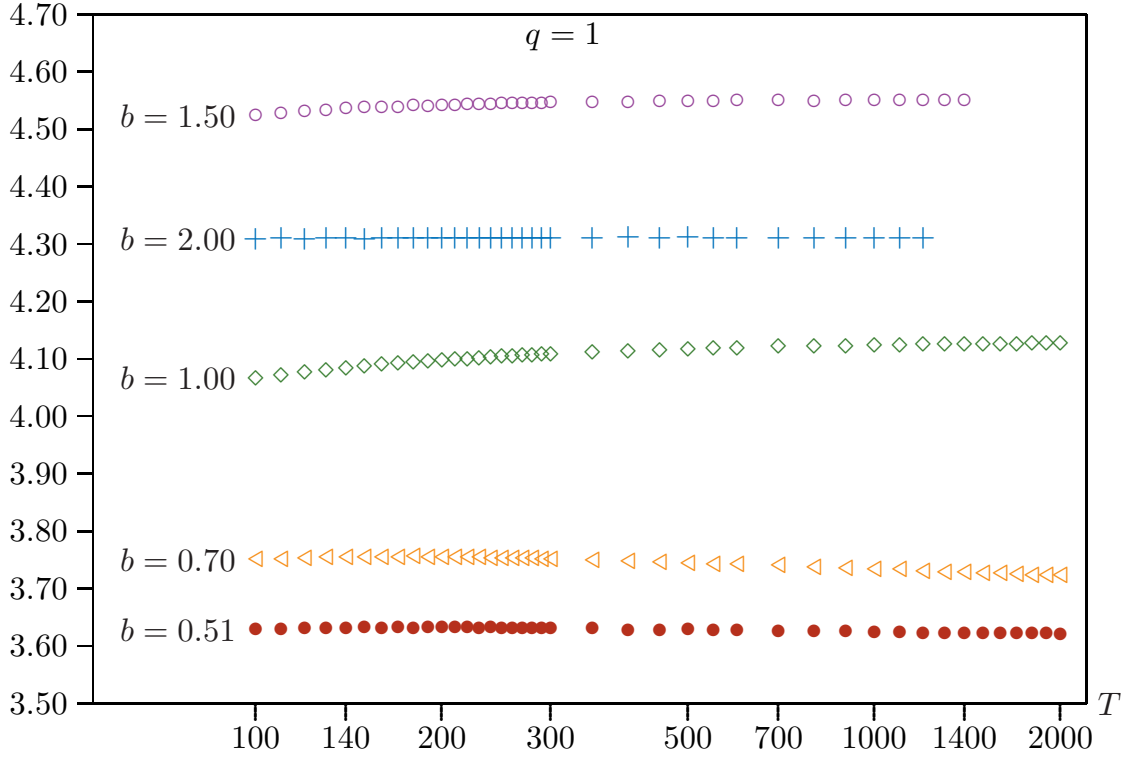


Figure 2: Estimated 0.95 quantiles as functions of T , RC case

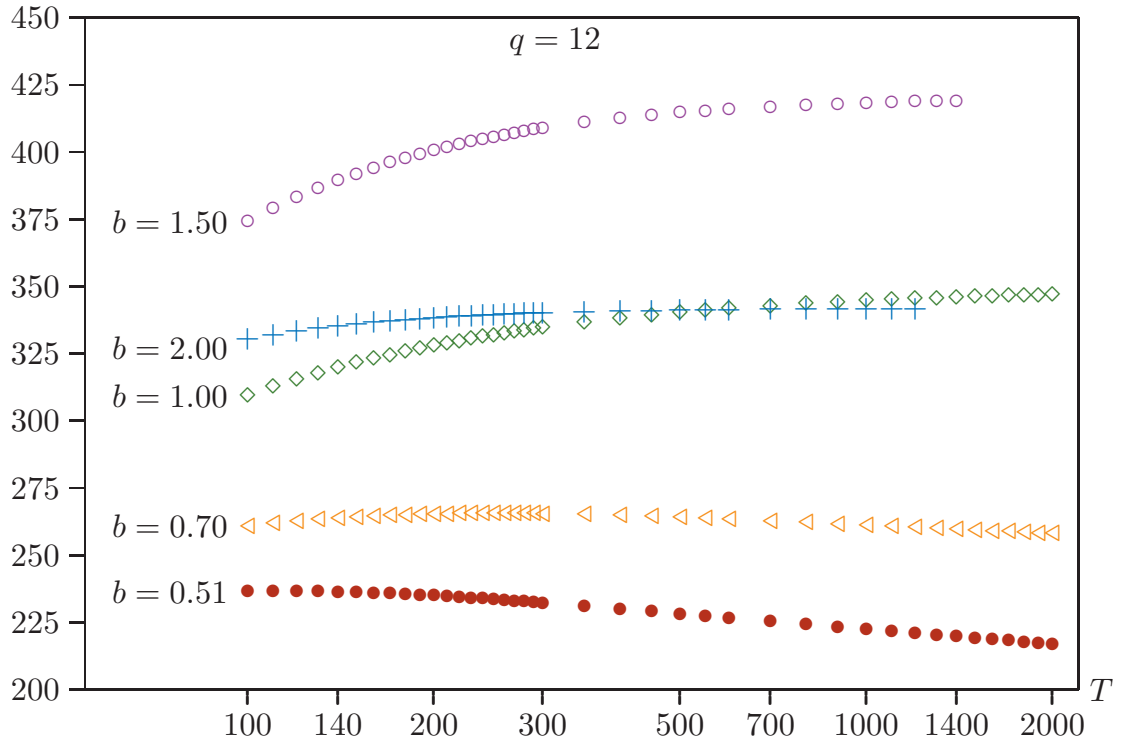
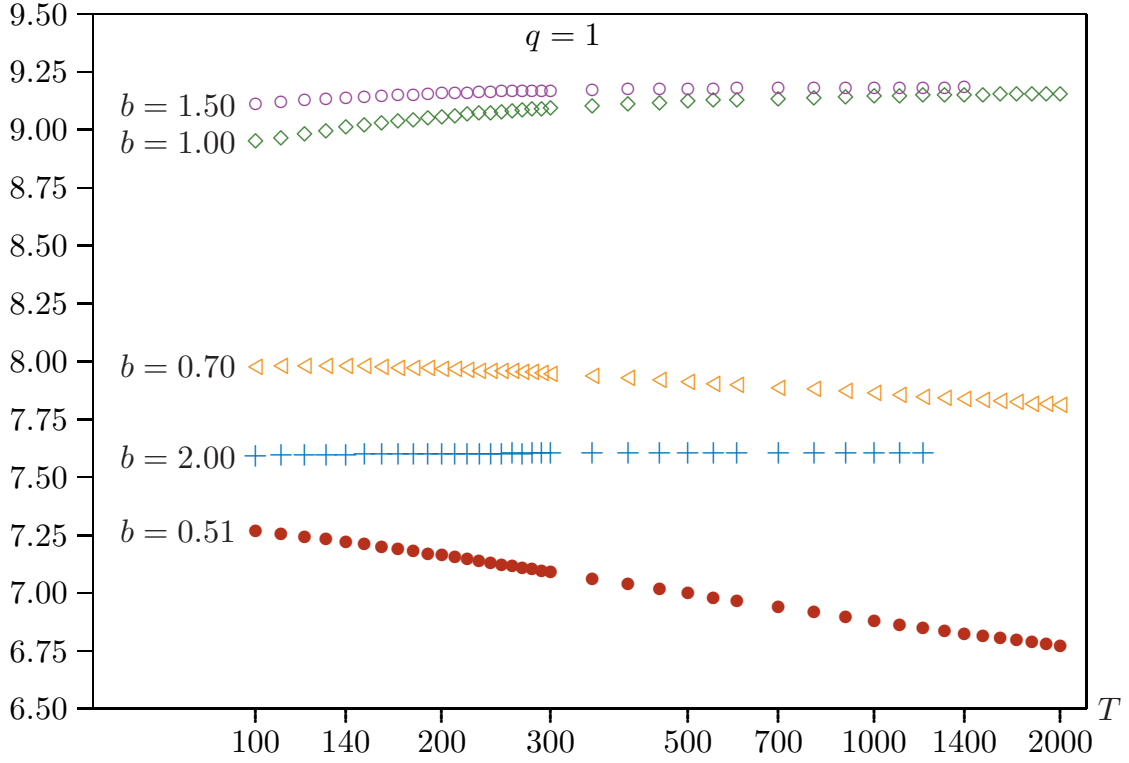


Figure 3: Estimated 0.95 asymptotic quantiles as functions of b , NC case

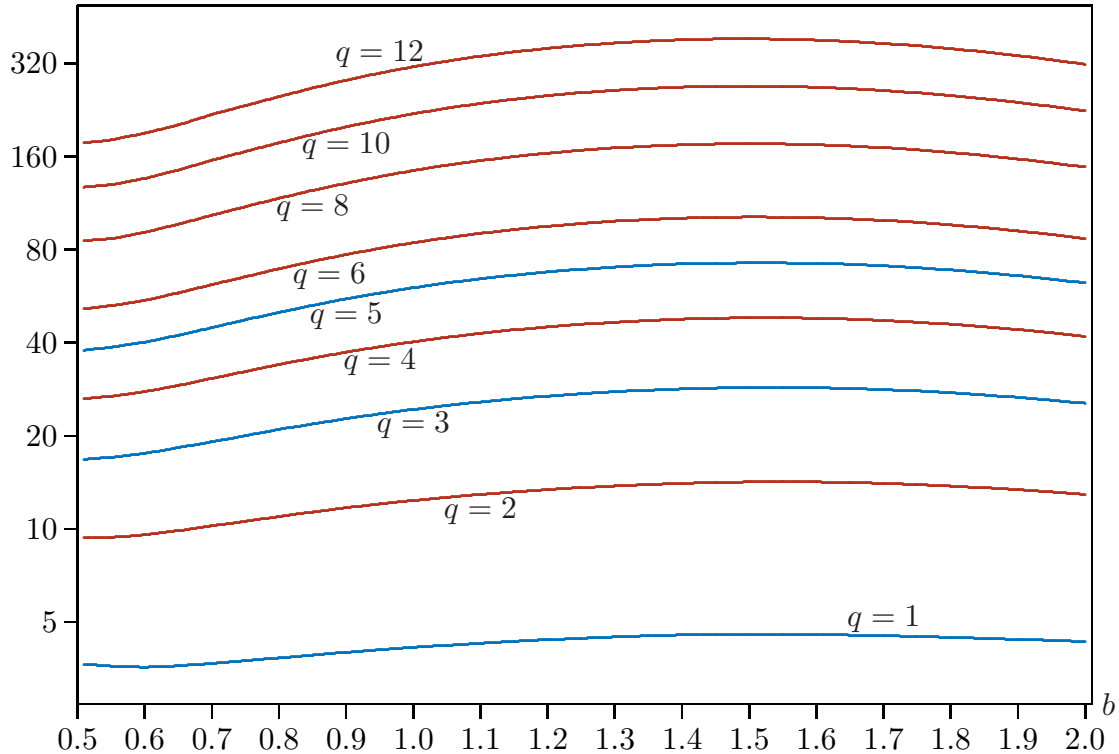


Figure 3 graphs asymptotic 0.05 critical values on a logarithmic scale as functions of b for most values of q in the NC case. It is evident that the asymptotic critical values are not monotonic in b . For all values of q , the critical value initially rises slowly as b increases (it actually drops slightly at first for $q = 1$), then rises quite rapidly for a while, achieves a maximum around $b = 1.5$, and declines quite substantially thereafter.

Figure 3 contains all the information that our experiments provide about asymptotic 0.05 critical values in the NC case, except for $q = 7, 9, 11$, which were omitted to avoid cluttering the figure. In principle, although it is not very convenient, one could simply read a desired critical value from the figure. It is clear from the figures that ignoring the presence of the parameter b would lead to substantial errors in inference, since quantiles depend in a non-trivial way on b . It would therefore be impractical to use tables of simulated critical values, which would extend to many pages. Moreover, such tables would provide less information than the critical values and P values provided by the computer programs which accompany this paper; see the next section.

4 Critical values and P values

The first step in estimating critical values and P values for a specified value \bar{b} is to calculate $\hat{F}_q^{\bar{b}}(\infty, \pi)$ using the $\hat{F}_q^b(\infty, \pi)$ for a number of values of b close to \bar{b} . As Figure 3 makes clear, the asymptotic distributions are smooth functions of b , so it should be possible to do this in a variety of ways.

We use the weighted local quadratic regression:

$$w(i)\hat{F}_q^{b(i)}(\infty, \pi) = \eta_1 w(i) + \eta_2 w(i)b(i) + \eta_3 w(i)b^2(i) + u(i), \quad (8)$$

where i runs from 1 to 31, $b(1) = 0.51$, $b(2) = 0.55$, and so on up to $b(31) = 2.00$, and $w(i) = \max(0, 1 - 5|b(i) - \bar{b}|)$. With this weighting scheme, only values of b near \bar{b} receive positive weights. For example, if $\bar{b} = 0.72$, there are eight such values:

$$0.55(0.15), 0.60(0.40), 0.65(0.65), 0.70(0.90), 0.75(0.85), 0.80(0.60), 0.85(0.35), 0.90(0.10).$$

Here the first number is $b(i)$ and the second is the weight $w(i)$. This case is typical, in that eight values of b receive positive weights. However, there may be as few as five when \bar{b} is close to 0.51 or 2.00. Our estimate $\hat{F}_q^{\bar{b}}(\infty, \pi)$ is simply $\hat{\eta}_1 + \hat{\eta}_2 \bar{b} + \hat{\eta}_3 \bar{b}^2$, where the $\hat{\eta}_j$ are the estimates from regression (8).

In order to obtain a P value for any observed test statistic $\text{LR}_T(q)$, a procedure for interpolating between the 221 values of $\hat{F}_q^{\bar{b}}(\infty, \pi)$ is needed. The procedure we use is a modified version of one proposed in MacKinnon (1996). Consider the cubic regression

$$G(\pi) = \gamma_0 + \gamma_1 \hat{F}(\pi) + \gamma_2 \hat{F}^2(\pi) + \gamma_3 \hat{F}^3(\pi) + e_\pi, \quad (9)$$

where $\hat{F}(\pi)$ denotes $\hat{F}_q^{\bar{b}}(\infty, \pi)$, and $G(\cdot)$ is an inverse CDF. The regressand is transformed in order to make the relationship between it and $\hat{F}(\pi)$ closer to linear. In MacKinnon (1996), which dealt with Dickey-Fuller tests, $G(\cdot)$ was the inverse standard normal CDF. That would be a bad choice here, because $F(\cdot)$ in no way resembles the inverse standard normal distribution. Instead, we note that if standard asymptotics applied, as when $b < 0.5$, then $\text{LR}_T(q) \xrightarrow{D} \chi^2(q^2)$. We therefore define $G(\pi)$ to be the inverse of the $\chi^2(q^2)$ CDF evaluated at π . The precise choice of $G(\cdot)$ does not matter very much.

It may seem a bit odd that the regressors in (9) are stochastic and the regressand is not. However, because the estimated quantiles are generally very accurate, the errors in variables bias that this induces is very small. This regression is estimated using only a small number of points, generally 9, in the neighborhood of the observed test statistic $\text{LR}_T(q)$. We first find the value of $\hat{F}(\pi)$ that is closest to $\text{LR}_T(q)$ and use the data for that value of π and the four values on either side of it. Slightly different procedures are used if the initial value of π is less than 0.002 or greater than 0.998. The estimated P value is simply

$$\hat{P} = 1 - G^{-1}(\hat{\gamma}_0 + \hat{\gamma}_1 \text{LR}_T(q) + \hat{\gamma}_2 \text{LR}_T^2(q) + \hat{\gamma}_3 \text{LR}_T^3(q)),$$

where the $\hat{\gamma}_i$ are the parameter estimates.

The $\hat{F}_q^{\bar{b}}(\infty, \pi)$ already give us asymptotic critical values for tests at any conventional significance level. However, we can obtain ones for tests at any level at all that should be slightly more accurate by turning regression (9) around:

$$\hat{F}(\pi) = \delta_0 + \delta_1 G(\pi) + \delta_2 G^2(\pi) + \delta_3 G^3(\pi) + e'_\pi. \quad (10)$$

Of course, this is not actually the inverse of equation (9), but it involves the same sort of approximation. As before, regression (10) is (usually) estimated for 9 observations around the desired level of the test. The estimated critical value for a test at level α is then simply

$$\hat{\delta}_0 + \hat{\delta}_1 G(1 - \alpha) + \hat{\delta}_2 (G(1 - \alpha))^2 + \hat{\delta}_3 (G(1 - \alpha))^3,$$

where the $\hat{\delta}_i$ are the parameter estimates.

The $\hat{F}_q^b(\infty, \pi)$ and a Fortran program that uses regressions (8), (9), and (10) to obtain estimated asymptotic P values and critical values are available from the authors' websites. The program is intended to be run interactively, but everything of any substance is done in subroutines that can easily be extracted and incorporated into larger programs.

5 Conclusion

We obtain asymptotic numerical distribution functions of likelihood ratio tests for fractional unit roots and cointegration rank. These can easily be used to obtain critical values and P values via a program that we make available. Because these distributions depend on a real-valued parameter, computational cost is very high. Moreover, the choice of model specification for the response surface regressions used to obtain the NDFs is more involved than is usual. We use model averaging rather than model selection to deal with uncertainty about the correct form of the response surface regressions.

6 References

- Buckland, S. T., Burnham, K. P., and Augustin, N. H. (1997). Model selection: An integral part of inference. *Biometrics* **53**, 603–618.
- Cavaliere, G., Rahbek, A., and Taylor, A. M. R. (2012). Bootstrap determination of the co-integration rank in VAR models. *Econometrica* **80**, 1721–1740.
- Davidson, J., and Hashimzade, N. (2009). Type I and type II fractional Brownian motions: A reconsideration. *Computational Statistics & Data Analysis* **53**, 2089–2106.
- Hansen, B. E. (2007). Least squares model averaging. *Econometrica* **75**, 1175–1189.
- Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control* **12**, 231–254.
- Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica* **59**, 1551–1580.
- Johansen, S. (2008). A representation theory for a class of vector autoregressive models for fractional processes. *Econometric Theory* **24**, 651–676.
- Johansen, S. and Nielsen, M. Ø. (2010). Likelihood inference for a nonstationary fractional autoregressive model. *Journal of Econometrics* **158**, 51–66.
- Johansen, S. and Nielsen, M. Ø. (2012). Likelihood inference for a fractionally cointegrated vector autoregressive model. *Econometrica*, forthcoming.
- L'Ecuyer, P., Blouin, F., and Couture, R. (1993). A search for good multiple recursive random number generators. *ACM Transactions on Modeling and Computer Simulation* **3**, 87–98.
- MacKinnon, J. G. (1996). Numerical distribution functions for unit root and cointegration tests. *Journal of Applied Econometrics* **11**, 601–618.
- MacKinnon, J. G., Haug, A. A., and Michelis, L. (1999). Numerical distribution functions of likelihood ratio tests for cointegration. *Journal of Applied Econometrics* **14**, 563–577.
- Nielsen, M. Ø. and Morin L. (2012). FCVARmodel.m: A Matlab software package for estimation and testing in the fractionally cointegrated VAR model. QED working paper 1273, Queen's University.
- Park, J. Y. (2003). Bootstrap unit root tests. *Econometrica* **71**, 1845–1895.